



LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS
partículas e tecnologia

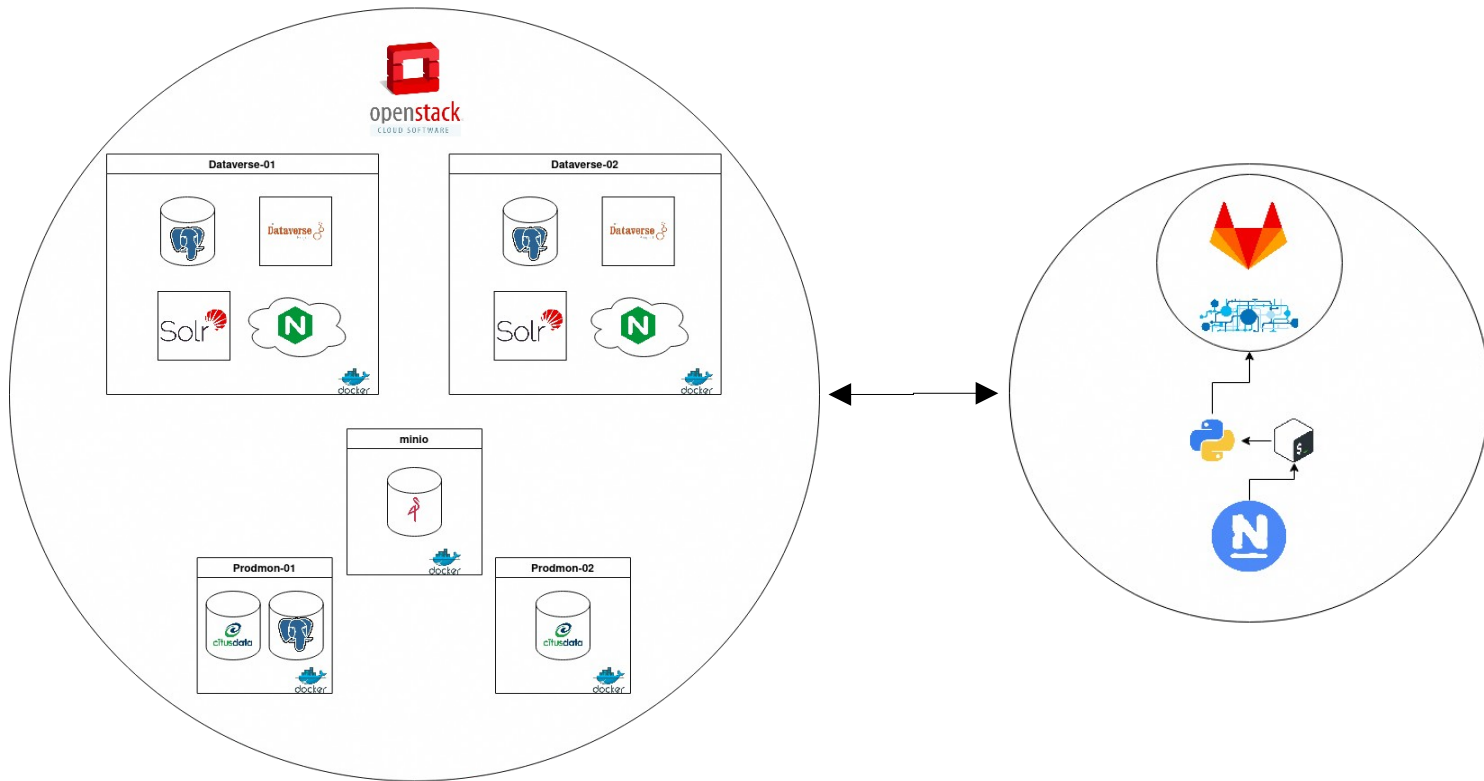
Dataverse benchmarking using opensource tools.

Zacarias Benta - LIP
César Ferreira - INCD
Fernando Ribeiro - FCCN
Jorge Gomes - LIP
João Martins - LIP
Mário David - LIP

Our challenges.

- Dataverse Service for National Scientific Computation Foundation;
- Fault tolerant service;
- No single point of failure;
- High availability.

Dataverse Distributed Architecture



Let's stress it out.

- No testing framework;
- No previous benchmark definition;
- Considered using the API;
- Jmeter to the rescue.

Jmeter FTW.

- Jmeter as a proxy;
- Non repeatable recorded events;
- API to the rescue;
 - Created 100 users and keys;
 - Used endpoints for interaction.
- Multi host/user requests.
 - 5 machines cluster;
 - Severall runs;
 - Increased nr of users/requests.

Jmeter FTW.

The screenshot displays the Apache JMeter 5.6.2 interface. The title bar shows the file path: `100_users_10_loop_file.jmx (/home/zbenta/Documents/DATAVERSE-FCC...e_dv_ds_file/100_users_10_loop_file.jmx) - Apache JMeter (5.6.2)`. The menu bar includes **File**, **Edit**, **Search**, **Run**, **Options**, **Tools**, and **Help**. The toolbar contains icons for file operations, test execution, and help.

The left sidebar shows a test plan tree for **Dataverse Python API Connection Tests**. The selected **Thread Group** contains the following elements:

- Uniform Random Timer
- Users DATAVERSE
- BeanShell Sampler
- Recording Controller
 - Knock on front door `${_threadNum}`
 - create_dataverse
 - Constant Timer
 - create_dataset
 - Constant Timer
 - upload_dataset_api_call
 - Constant Timer
 - delete_dataset
 - Constant Timer
 - delete_dataverse
 - Constant Timer
- View Results Tree
- View Results in Table
- Debug Sampler
- HTTP(S) Test Script Recorder

The right sidebar shows the **Thread Group** configuration panel:

- Name:** Thread Group
- Comments:** (empty)
- Action to be taken after a Sampler error:** Continue Start Next Thread Loop Stop Thread Stop Test Stop Test Now
- Thread Properties:**
 - Number of Threads (users):** 100
 - Ramp-up period (seconds):** 1
 - Loop Count:** Infinite 10
 - Same user on each iteration
 - Delay Thread creation until needed
 - Specify Thread lifetime
 - Duration (seconds):** (empty)
 - Startup delay (seconds):** (empty)

Jmeter FTW.

The screenshot displays the Apache JMeter 5.6.2 interface. The title bar shows the file path: `100_users_10_loop_file.jmx (/home/zbenta/Documents/DATAVERSE-FCCN/TESTES/PROD/Docker_Tests/100_10_loops_create_delete_dv_ds_file/100_users_10_loop_file.jmx) - Apache JMeter (5.6.2)`. The menu bar includes File, Edit, Search, Run, Options, Tools, and Help. The toolbar contains various icons for file operations and test execution. The left sidebar shows a tree view of the test plan, with the following structure:

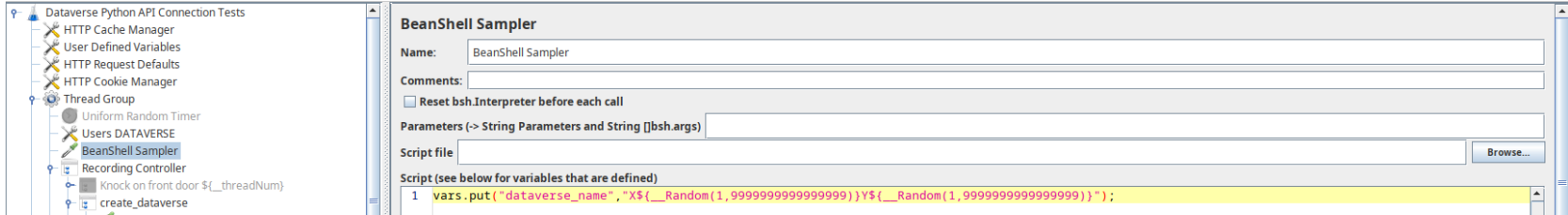
- Dataverse Python API Connection Tests
 - HTTP Cache Manager
 - User Defined Variables
 - HTTP Request Defaults
 - HTTP Cookie Manager
 - Thread Group
 - Uniform Random Timer
 - Users DATAVERSE
 - BeanShell Sampler
 - Recording Controller
 - Knock on front door \${_threadNum}
 - create_dataverse
 - create_dataverse-operation \${_threadNum}
 - HTTP Header Manager
 - Response Assertion
 - Constant Timer
 - create_dataset
 - Constant Timer
 - upload_dataset_api_call
 - Constant Timer
 - delete_dataset
 - Constant Timer
 - delete_dataverse
 - Constant Timer
 - View Results Tree
 - View Results in Table
 - Debug Sampler
 - HTTP(S) Test Script Recorder
 - View Results Tree

The main window displays the configuration for the selected `create_dataverse-operation ${_threadNum}` element, which is an **HTTP Request**. The configuration is as follows:

- Name:** `create_dataverse-operation ${_threadNum}`
- Comments:** (empty)
- Basic Tab:**
 - Web Server:**
 - Protocol [http]: (empty)
 - Server Name or IP: (empty)
 - Port Number: (empty)
 - HTTP Request:**
 - Method: `POST`
 - Path: `/api/dataverses/polen`
 - Content encoding: `UTF-8`
 - Options: Redirect Automatically, Follow Redirects, Use KeepAlive, Use multipart/form-data, Browser-compatible headers
- Parameters Tab:**
 - Body Data:

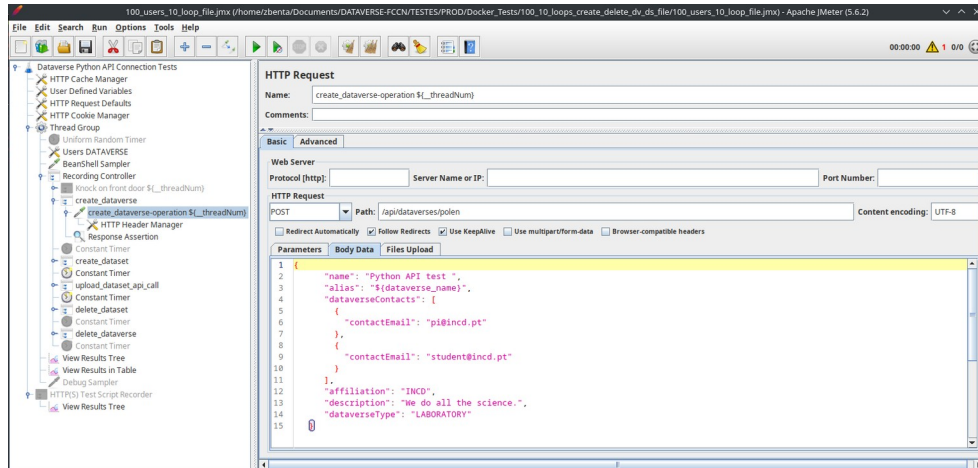
```
1 {
2   "name": "Python API test ",
3   "alias": "${dataset_name}",
4   "dataverseContacts": [
5     {
6       "contactEmail": "pi@incd.pt"
7     },
8     {
9       "contactEmail": "student@incd.pt"
10    }
11  ],
12  "affiliation": "INCD",
13  "description": "We do all the science.",
14  "dataverseType": "LABORATORY"
15 }
```

Jmeter FTW.



The screenshot shows the configuration for a BeanShell Sampler in JMeter. The left sidebar displays a test plan structure with the following elements:

- Dataverse Python API Connection Tests
 - HTTP Cache Manager
 - User Defined Variables
 - HTTP Request Defaults
 - HTTP Cookie Manager
 - Thread Group
 - Uniform Random Timer
 - Users DATVERSE
 - BeanShell Sampler
 - Recording Controller
 - Knock on front door \${__threadNum}
 - create_dataverse



The screenshot shows the configuration for an HTTP Request in JMeter. The left sidebar displays a test plan structure with the following elements:

- Dataverse Python API Connection Tests
 - HTTP Cache Manager
 - User Defined Variables
 - HTTP Request Defaults
 - HTTP Cookie Manager
 - Thread Group
 - Uniform Random Timer
 - Users DATVERSE
 - BeanShell Sampler
 - Recording Controller
 - Knock on front door \${__threadNum}
 - create_dataverse
 - create_dataverse-operation \${__threadNum}
 - HTTP Header Manager
 - Response Assertion

HTTP Request

Name: create_dataverse-operation \${__threadNum}

Comments:

Basic Advanced

Web Server

Protocol [http] Server Name or IP: Port Number:

HTTP Request

POST Path: /api/dataverse/pcpoln Content encoding: UTF-8

Redirect Automatically Follow Redirects Use KeepAlive Use multipart/form-data Browser-compatible headers

Parameters Body Data Files Upload

```
1 {
2   "name": "Python API test ",
3   "alias": "${dataverse_name}",
4   "dataverseContacts": [
5     {
6       "contactEmail": "pi@incd.pt"
7     },
8     {
9       "contactEmail": "student@incd.pt"
10    }
11  ],
12  "affiliation": "INCD",
13  "description": "We do all the science.",
14  "dataverseType": "LABORATORY"
15 }
```


Jmeter FTW.

The screenshot displays the Apache JMeter 5.6.2 interface. The title bar shows the file path: `100_users_10_loop_file.jmx (/home/zbenta/Documents/DATAVERSE-FCCN/TESTES/PROD/Docker_Tests/100_10_loops_create_delete_dv_ds_file/100_users_10_loop_file.jmx) - Apache JMeter (5.6.2)`. The menu bar includes File, Edit, Search, Run, Options, Tools, and Help. The toolbar contains various icons for file operations and execution. The status bar at the bottom right shows `00:00:00`, a warning icon, and `1 0/0`.

The left sidebar shows a tree view of the test plan. The selected element is `Users DATAVERSE`, which is a `Thread Group` containing several samplers and controllers, including `Uniform Random Timer`, `Beanshell Sampler`, `Recording Controller`, `Knock on front door`, `create_dataverse`, `create_dataverse-operation`, `HTTP Header Manager`, `Response Assertion`, `Constant Timer`, `create_dataset`, `upload_dataset_api_call`, `delete_dataset`, `delete_dataverse`, `View Results Tree`, `View Results in Table`, `Debug Sampler`, and `HTTP(S) Test Script Recorder`.

The main panel displays the configuration for the selected `Users DATAVERSE` element, titled `CSV Data Set Config`. The configuration includes:

- Name:** Users DATAVERSE
- Comments:** (empty)
- Configure the CSV Data Source:**
 - Filename:** `/home/zbenta/Documents/DATAVERSE-FCCN/Desenvolvimento/qty/user_data_prod_newdb.csv`
 - File encoding:** (dropdown menu)
 - Variable Names (comma-delimited):** `UserID,UserName,UserApiKey`
- Ignore first line (only used if Variable Names is not empty):** `True`
- Delimiter (use '\t' for tab):** `,`
- Allow quoted data?:** `False`
- Recycle on EOF?:** `True`
- Stop thread on EOF?:** `False`
- Sharing mode:** `All threads`

Jmeter FTW.

The screenshot displays the Apache JMeter 5.6.2 interface. The title bar shows the file path: `100_users_10_loop_file.jmx (/home/zbenta/Documents/DATAVERSE-FCCN/TESTES/PROD/Docker_Tests/100_10_loops_create_delete_dv_ds_file/100_users_10_loop_file.jmx) - Apache JMeter (5.6.2)`. The menu bar includes File, Edit, Search, Run, Options, Tools, and Help. The toolbar contains various icons for file operations and test execution. The status bar at the bottom right shows a timer at `00:00:00`, a warning icon, and `1 0/0`.

The left sidebar shows a test plan tree for "Dataverse Python API Connection Tests". The selected element is "HTTP Header Manager" under the "create_dataverse-operation" sampler. The main panel displays the configuration for the "HTTP Header Manager":

HTTP Header Manager

Name: HTTP Header Manager

Comments:

| Headers Stored in the Header Manager | |
|--------------------------------------|--|
| Name | Value |
| Content-Type | application/json |
| X-Dataverse-key | \${UserApiKey} |
| Accept-Encoding | gzip, deflate, br |
| User-Agent | python-requests/2.28.1 |
| Accept | */* |
| dataverse_name_1 | X\${_Random(1,9999999999999999)}Y\${_Random(1,9999999999999999)} |

Our issues.

- Slow response times;
- Bottleneck hard to discover;
- No prior detective work experience.

The tools we used for debug.

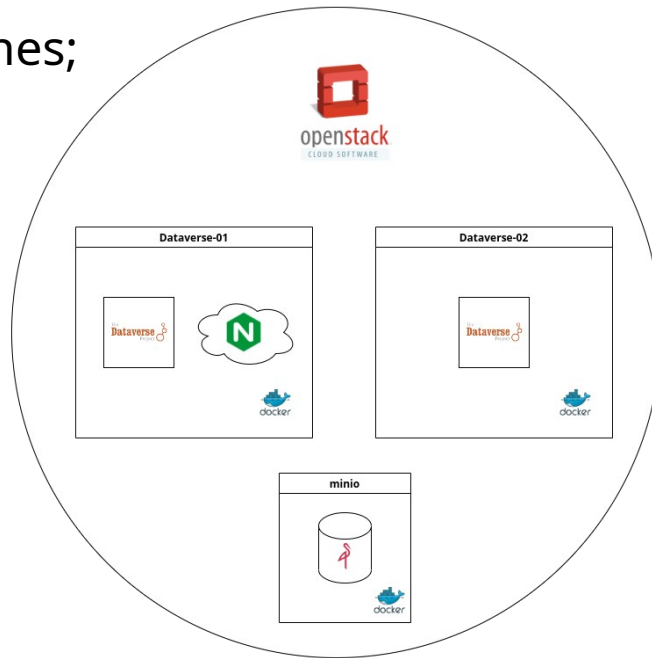
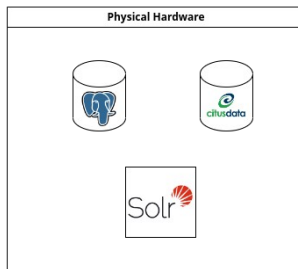
- ps;
- htop;
- docker stats;
- nethogs;
- dd;
- iostats;
- pgbench.

Our findings.

- I/O throughput, too slow;
 - Saving files took a lot of time;
 - Database performance was affected.
- Onions will eventually make you cry.

Our solution.

- Create a new architecture;
- Moving services to physical machines;
- Use nvme for faster performance.



Performance comparison.

Before

| Requests | Executions | | | Response Times (ms) | | | | | | | Throughput | Network (KB/sec) | |
|----------|------------|------|---------|---------------------|-----|--------|----------|-----------|-----------|-----------|----------------|------------------|------|
| Label | #Samples | FAIL | Error % | Average | Min | Max | Median | 90th pct | 95th pct | 99th pct | Transactions/s | Received | Sent |
| Total | 282 | 0 | 0.00% | 54106.24 | 0 | 137482 | 43689.50 | 136813.40 | 137152.45 | 137437.42 | 0.91 | 0.57 | 0.51 |

After

| Requests | Executions | | | Response Times (ms) | | | | | | | Throughput | Network (KB/sec) | |
|----------|------------|------|---------|---------------------|-----|-------|---------|----------|----------|----------|----------------|------------------|-------|
| Label | #Samples | FAIL | Error % | Average | Min | Max | Median | 90th pct | 95th pct | 99th pct | Transactions/s | Received | Sent |
| Total | 5000 | 0 | 0.00% | 4474.00 | 0 | 11483 | 4006.50 | 9058.80 | 9435.90 | 10387.89 | 20.32 | 13.05 | 12.09 |



Thanks

Any questions?

You can find me at zacarias@lip.pt